“Data Something”: Assessing Tools, Services and Barriers for Research Data Collaboration at the University of Edinburgh [Report]

Report on a small-scale study carried out by Dr Tamar Israeli, Western Galilee College Library, as part of her work placement visit with the Research Data Support team in Library and University Collections, September/October 2019.

Contents

"[The University tools and services] all start with 'data something' and I need to close my eyes and think which is for what." **Dr Zielinski**

Executive Summary

We carried out a small-scale study in order to gain a better understanding of the tools researchers use to collaborate around data, and to explore the barriers and difficulties that prevent researchers from using institutional tools and services. Six semi-structured interviews were conducted with researchers from the University of Edinburgh, representing different schools, and all of whom collaborate with other researchers on a regular basis on either small- or large-scale projects.

We found that participants use many different tools, both institutional and commercial, to collaborate, share, analyse and transfer documents and data files. Decisions about which tools to use are based on data types, data size, usability, network effect and whether their collaborators are in the same institution and country. Researchers tend to use institutional tools only if they are very simple and user friendly, if there is a special requirement for this from funders or principal investigators (PIs), or if it is directly beneficial for them from a data analysis perspective. Sharing beyond the immediate collaboration is only a secondary concern. A major issue was the need for tools that answer projects' particular needs. Researchers are generally well aware of the need to keep their data where it will be safe and backed-up, and are not concerned about the risk of data loss.

Context and Methodology

Context: The purpose of the study was to gain a better understanding of the tools that researchers use to collaborate around data, and to explore the barriers and difficulties that prevent researchers from using institutional tools and services.

Methodology: semi-structured interviews were conducted with six researchers from the university. The interviews lasted between half an hour and an hour. They were recorded and then analysed to identify major themes. Participants were university researchers from different disciplines and faculties who collaborate with other researchers on a regular basis on either small- or large-scale projects.

Literature Review

When a team of two or more researchers starts collaborating, they need to choose how to store, share and edit their files. Many use cloud-based services such as Google Drive and Dropbox or simply distribute their files by email and store them on their personal computers. Studies have shown that teams often decide which sharing service to use based on what a majority of the collaborators had access to or were experienced with, rather than specific features (Capra, Vardell, & Brennan, 2014; Rader 2009).

Research data files differ from other information files. Data files' size are sometimes exceptionally large or need special software to access and analyses the data. Some research data are unique and cannot be replaced if destroyed or lost. Some are sensitive and cannot be transferred, whether locally or internationally, without suitable security in place. Good data management practices include data protection. Common drives offered by universities back up research products and keep them protected and usable. They are ideally trusted, and comply with funders' and other stakeholders' standards. However, studies show that researchers frequently rely on other methods

of storage such as flash drive, external hard drive, commercial Cloud storage or emails (e.g., Renwick, Winter & Gill, 2017; Alexogiannopoulos, McKenney, & Pickton, 2010).

A key question is whether there are particular requirements based on specific research fields. Indeed, previous studies have shown that there are disparities in the way data is curated and shared among and within field of studies (e.g., Tenopir, Sandusky, Allard, & Birch, 2014 ; Akers & Doty, 2013). To understand the different needs, when looking for participants, we turned to researchers from different disciplines and faculties. All the participants in this study were researchers who collaborate with other researchers on a regular basis on small or large scale projects.

The Interviews: What tools are researchers using to collaborate?

Researchers referred to the various tools they use for storing, transferring and editing documents and data. The decision to use specific tools is based on a number of factors, such as: Who are the collaborators (internal or external)? What size are the files? How sensitive is the data? What is the purpose of collaboration (e.g. writing a paper or analysing data)? Participants discussed commercial (often free-to-use) software as well as institutionally-provided tools.

Email

Email is a simple and well-known tool for transferring files which are not excessively large. Each user selects the software they find most convenient, and can add a context concerning the attached file. Email is found to be particularly advantageous when working with partners in different countries or institutions.

Professor Andrew Millar, Chair of Systems Biology and Business Owner of the University Research Data Service, said that email is always extensively used.

Dr Tomasz Zielinski, Computational Resource Manager at the School of Biological Sciences, believes that scientists use email because it is unfailingly reliable. Most data files in biology are in Excel format, and researchers usually provide only minimal descriptive information like headings inside those files. Other typical files are in MS Word format, and are also transferred by email. According to Zielinski, collaborators use emails to notify colleagues of any changes because it does not require additional setup, also not all of the tools are available when collaborating with people outside the university.

Holly Tibble, a PhD student in Medical Informatics, also mentioned problems she has encountered while working with partners in different countries. She uses email for transferring small data files which do not require additional security. While she was working on one of her projects with 26 groups around the world, it was forbidden to transfer data. As a way around this problem, she sent only the scripts to her colleagues so they could analyse their own data and then return their results to her. They used email to exchange these files.

Wenlong Li, a PhD student in IT law, is working on a study with three Chinese colleagues. One of them is based in a different institution in the UK, and one is located in China. Recently they wrote an abstract for a conference and used email to transfer their files. Email helps them to overcome another barrier, namely the difficulty of finding a particular instance of a word on Google Docs files which contain comments. As it is sometimes difficult to track changes on Google Docs, they use email to inform partners about changes.

Dr Brian Cahill, Research Programme Manager, said that people send their documents to him by email as this is often a very convenient way for partners to follow the development of the versions

when not all the researchers are from the same institution. They use Word files with tracked changes for their shared documents.

Google Drive, Google Docs, Dropbox, WeChat

Wenlong Li's field of study is law. As he deals with public data there are no security or privacy issues. He and his colleagues cooperate and share documents using Google Docs, which he finds to be a very efficient tool. They use it to edit files and then send a link to the file location. Another tool he uses to transfer files and communicate with his Chinese colleagues is WeChat, which can be described as the Chinese equivalent of Facebook.

Dr Cahill said that while he himself does not use Google Docs, he knows other people who do use it to cooperate. He thinks that email is probably not the most efficient tool for dealing with versions; Google Docs is probably more efficient and he has worked in institutions where they had their own versions of Google Docs for study proposals. Their product was less user friendly but some institutions have a policy of not using Google products.

Dr Cyril Pernet, a cognitive neuroscience researcher, said that he mostly uses Google Docs for papers. As it is integrated with Zotero he can manage his bibliography there. Google Docs has a version history which makes it more convenient than Word files. He uses Overleaf with some of his colleagues because it simplifies collaboration when working with LaTex. Nevertheless, many people prefer not to work with it.

Dr Pernet uses *DataSync* when data need to be secured, but mentioned that this does not concern some of his colleagues who simply use Google Drive and Dropbox to transfer their data. He also uses Google Drive when the data is anonymised. He said that he and his colleagues buy extra space in Google Drive, paying for it themselves.

Professor Millar uses tools like Google Drive and Dropbox especially when collaborating with external colleagues. Although he can use *DataSync* to send links to files, his colleagues prefer to receive them on interfaces that they are familiar with. Some researchers are very disciplined and very structured when using Google Drive and Dropbox, he said. The problem is that when researchers exceed their space limits on these tools they remove some of the data to make room for other data, which possibly makes this tool a less secure option for data storing. On second thoughts, he said, institutional tools actually have the same problem. When working with external collaborators one has to work with the tools that they use. External collaborators who work with university tools and services must open an external account, and since they collaborate on different projects with different institutions they prefer to use their Google Drive and Dropbox accounts.

Wiki

Dr Zielinski thinks that *Wiki* is a very simple and flexible tool for describing and sharing the data which is mostly used for collaboration. It is used to organize the data connected to a particular topic, for example data files related to paper under development or meeting notes and reports, or for small sized documents like common protocols. He also mentioned that in many research data management plans, *wiki* is declared as the tool that are going to be used since it is simple, intuitive, easy to use and there are no barriers to setting up an account. Dr Zielinski described his experience a large project on biological clocks for which a special tool for collecting and analysing data was built but the collaborators still used a wiki for everything that did not require sophisticated analysis.

Professor Millar and his colleagues encouraged their partners in previous projects to use more structured tools like *wikis*. They used Edinburgh University's Confluence Wiki as a tool for storing small data and manuscript files. Twelve years ago, before the wiki was provided by the University,

the CSBE research centre built its own wiki where they could create codes for maintenance and for managing staff reports. These functions were automated, unlike the University's wiki. They also used it to exchange some research data files among collaborators. Millar explained that when they are engaged in a long-term study, researchers are willing to open external accounts at Edinburgh University for using the wiki and to learn how to use the wiki's platform which is different from the one they use in their own institutions.

Microsoft Office 365

Microsoft's *SharePoint* was also mentioned as a simple common drive for storing and sharing information. There is no need for version management, as it is secured and works well with other Microsoft products.

Dr Cahill mentioned using *SharePoint* for administrative needs, describing it as a simple tool that offers secure and easy access to personal data. He also considers it as a simple common drive that can be used for data sharing in small teams where collaboration is less sophisticated.

Dr Zielinski considers SharePoint to be a perfect tool for sharing and simultaneously working on documents. It can usually be easily accessed from any of his devices and also supports versioning, so is an excellent tool for collaborative writing papers or reports. Unfortunately, it seems that not everyone finds it easy to use or access. His supervisor, for example, has to log in each time he needs access and therefore prefers transferring files via mail.

Professor Millar has so far not used Microsoft tools like SharePoint and OneDrive in his research projects. He understands that the University's provision of these tools was intended to support administrative work, rather than data, but some of his lab use them for manuscript preparation.

Institutional Networked Storage

Wenlong Li was pleased with the fact that the university has provided him with 20 GB free space on the server free of charge and that he can sync his files and access them from different devices and places. However, he only uses it for his one note files which are better supported on the Microsoft system. He prefers to use Google Drive for other documents.

Dr Zielinski thinks DataStore (the institutional drive) is an excellent service for active data. People in typical cases are not required to pay for storing their data and they can easily collaborate with colleagues from Edinburgh University. However, if another partner from a different institution joins the project, a different solution is needed. When collaborating with partners from different institutions, researchers tend to move to Dropbox or Google Drive. There is no alternative way for partners in Newcastle, for example, to access the files, he said. Dropbox is satisfactory and it is secured. Dr Zielinski is not aware of any policy that bans its use. If it becomes illegal in the future, the university will have to provide practical tools for sharing.

Dr Pernet said that when he collaborates with colleagues from the University they use *DataStore* or *DataShare* (if the data was already published). This is not the case when he collaborates with external partners.

Specific Issues

Data Management Planning

Participants discussed factors to be considered when writing a data management plan.

Holly Tibble is a statistician. When writing a data management plan she considers data size and how to analyse it. Working with *eDRIS*, a secure tool for working with sensitive data doesn't really give her a choice. The data is transferred via the National Safe Haven contingent upon committee approval. With regard to other files, she and her partners decide where to store them based on the format of the file the study will produce.

Dr Cahill had just launched a new project and was still in the process of recruiting researchers and had thus not yet formulated a data plan. His researchers are required to undergo training in data ethics, and learn how to submit applications to ethics committees in different countries, which can be quite problematic as each group has to get approval from its own ethics committee. He said that he did not intend to dictate to his groups on how to share data, but did suggest on a webinar he delivered at an earlier stage, that they include data management plans in their proposals as this may help them in the review process. He was not sure if Horizon 2020 requires a data management plan, but is very supportive of the idea because he believes that it facilitates collaboration and it is good to have a structure which you can work with.

According to Dr ZielinskI, sadly, there is usually not much data sharing on large, multi-partners projects, only fund sharing. Groups tend to do their own science with limited data exchange. If they are required to declare which tool they would use for collaboration they would probably choose to use a wiki because of its simplicity.

Dr Pernet teaches some courses on data management. He tries to persuade his students and partners to organize all their project files in one place, but this is difficult, especially with PIs who are accustomed to managing their information in a certain way and don't like to take advice on how to do things. Some researchers keep some of their files on their laptops and some on the server. Pernet said they behave as if the data belong to them, while in fact they belong to the university and to the funder so they cannot do whatever they like with.

Data Security and Privacy

Tibble works with Scotland's national records which are very sensitive. All data access issues are taken care of by a *Data Safe Haven.* She may use her laptop for datasets which are not sensitive, but is legally bound to prevent anyone from viewing her computer while she works. She keeps high-level authentication with passwords and codes for data on her laptop and other devices. When asked about data loss she said that everything is backed up on institutional or external drives and even if she would lose her laptop or external drive, her team in *GitHub* would be able to send her the data within a day. Holly was not concerned with data anonymity. She said that the files are generally anonymised. They usually don't contain names, but special identifiers called CHI numbers.

Li is not concerned with data loss either. He stores and shares file using the university server, email or Google Docs. He understands that using Google's tool has disadvantages in terms of privacy. He said that if for some reason he did not want Google to track him and use his data, he would compare tools and would probably choose the university server which is more secure and stable.

Dr Cahill believes that the university should provide enough space for researchers so they that they do not have to store their data on external drives which, like all mechanical tools, are wont to crash eventually. He also thinks that the data should be stored on the university server in case one of the researchers leaves the group or institution, in which case finding the data would be very difficult. He said that after five years it is impossible to find data files that are stored on external drives. Dr Cahill recalls losing the date from half a year's work, and thinks that researchers should be aware of that possibility and make sure to back up their data, preferably on a cloud system. Of course, the cloud

provider can go bankrupt as well; there will be no perfect solution for the next hundred years, but it is preferable to USB sticks, external drives or laptops. Cahill believes that the university should find a way to ensure that information is properly secured and ascertain that researchers follow the correct protocol for securing their data, and if necessary, compel them to do so.

Dr Zielinski is sure that most researchers have problems with data of which they are simply not aware.. Although there are often backup policies in place, the actual backups are not checked before they are needed. He doubts that for example CD backups may be still read after a few years. He is not sure that the data would be accessible if someone were to request it. In his opinion the funders' ten-year requirement for storing data is unrealistic. Scientists generally use new data. Only the institution can hold data for ten years because the research groups and their budget operates on shorter time scale. Dr Zielinski argues if someone doesn't touch their data for three years, then these data areeither not important or no longer useful. Besides, how can one find data after three years? If someone has an Excel file in his lab, how could other people know about it? It would be realistic to focus on ten percent of the data, then scientists might make an effort to store and properly described their most important files and present them in a well-organized fashion.

Dr Pernet keeps all his files and data on the server or on Google Drive's cloud so he is not concerned with data loss but is aware of security issues. Even when taking steps to anonymize data it remains possible to track a patient's identity because of the screened images. The question is how far one can go with it.

Until now, Professor Millar's groups have chosen not to work with data that need anonymization. In the past there were no experts to help with this process. The situation has now changed and they are considering working with medical data. He and his colleagues are well aware of possible data loss and invest much effort in prevention.

Big Data Files

Data files are sometimes too large for storing and sharing on the services mentioned above. Holly Tibble said that when her files are too big for email and don't require security she transfers them using WeTransfer. Dr Pernet said that he doesn't generally share data from his projects that involve brain imaging. The data files are huge so they are rarely shared. When he does share he uses *DataSync.* He says that it is as easy to use as Dropbox. He only needs to send the link to the data. Other collaborators can download the data and it secured.

Pernet is promoting the Open Science Framework where people can connect all their project files and products from many different competing tools like Google Drive, Dropbox, GitHub, DataShare and DataStore.

Professor Millar mentioned using Zenodo as a generic repository for public purposes but said it was not as useful as customised resources with added-value tools, because it is just a bucket of files. There are currently no available integrated tools, though other repositories may be linked.

Recommendations: Barriers to using institutional tools and services

1. Tools must be more user friendly

All the participants agreed that tools must be simple and user friendly in order for people to use them. Researchers will probably use the simplest tool they can unless forced to do otherwise.

Dr Zielinski said that a tool with multiple features tends to confuse, and people look for something simpler. He is fairly sure that sharing is a secondary purpose for most researchers and that they use

a tool only if it benefits their research. Researchers will not invest time in uploading their data files in a neat and organized way unless they have been forced to do it, and then they will probably not remember all the necessary details to create extended metadata. Although he believes DataShare is a good tool for public access, he thinks that it is important to have a simple automatic way to transfer data from other tools to DataShare. If there are only a few files people simply copy and paste, but it should also be a way to deal effectively with a larger quantity of files. In most of his projects, the custom tools, which were developed for data sharing, addressed a narrow, specific research need. Data that did not directly benefit from the custom tools tend to end up in simpler systems like the University wiki.

Professor Millar also spoke about added value. It is very challenging to accustom investigators to working in a certain collaborating routine, and if this is to succeed, one must offer them an immediate added value such as the option to connect to other tools like *Jupyter* that enable data analysis.

Holly Tibble deals with sensitive data. She keeps her data protected with eDRIS National Safe Haven and uses a high level of authentication for data that is stored on her laptop. She expressed great concern about security issues and took a GPPR course to learn how to deal with them, but she is still anxious and wants to be on the safe side. She is sure that no one can steal her data but it takes her about half an hour to access them which is quite exhausting. Her two supervisors do not have access to the data as they didn't take the test needed for that. They simply trust her to do the right thing. When talking about versions she noted that when she uses the National Safe Haven she doesn't have access to the internet so she can't use *GitHub* which is her preferred tool for version control.

Wenlong Li explained that he uses Google Drive even though he is concerned about his privacy because Google Drive is user friendly, faster and has a better interface. He said Google also provides additional features that support his work. There was also the network effect. All his friends use it so it is easy for him to use it as well. He said that the tools the university provides are less accessible than the commercial tools. He would consider moving to the university tools if privacy issues arise. Right now it demands too much cognitive effort.

Dr Cahill agreed that that the tools provided by the University should be easy and comfortable to use. He thinks that the data analysis system within the bigger system must be good. If this system is difficult and unfriendly, everyone will reject it. It is very difficult to build a friendly system but the teams themselves can do it. Systems which require one to search for one's files or copy and paste them to other systems, present a problem. It takes time to develop a simple and easy system to analyse and present data successfully. Dr Cahill recalled a former project in Germany where they invested money in electronic lab notebooks but, as the software was in German and not user friendly, they hardly used it.

There was a general consensus among the participants that only programmers will use version control systems like GitHub or *Subversion.* It is too complicated even though there are courses for learning it. This system also doesn't work on documents like Excel and Word because they are binary. One cannot see the changes. Dr Pernet mentioned another tool for version control - HackMD which is like GitHub but easier and is good for note-taking during the project. This tool is based on mark-up. One doesn't have to be a programmer to use it, nor does one have to know the basic mark-up commands because it has a graphic user interface that one can use. Perent himself is a programmer and uses GitHub for managing parts of the projects which are more sophisticated than notes. Prof Millar mentioned some tools for version controls which are part of a larger system and are suitable for people who are not programmers. Of course these tools offer fewer options.

There is no problem with documents' versions with tools like Wiki, SharePoint or Google Docs, and as Dr Zielinski and Holly Tibble pointed out, there are no version problems with data because there are no new versions, only new data. Dr Zielinski referred to tools like *DataVault* as being useful because of the requirement to store data for ten years but thought that many researchers avoid using it because it is expensive and the upload process is complicated.

Tibble also mentioned the stability issue. She and her partners are using National Safe Haven for a project with very sensitive data which uses Scotland national records. They are very strict with their data security. She said that one of the problems with this data safe haven is that memory issues arise when several people work simultaneously, even if they are not from the same project. Occasionally there is no access to the data because the system crashes or because of maintenance work.

When I asked Dr Pernet why he and his colleagues pay for Google Drive space with their own money when they can have university storage space for free, he explained that when he works from home or from abroad it is easier to access Google Drive. He needs VPN for the institutional drive and it takes a couple of steps to reach his files.

Dr Pernet referred to Microsoft products: "No one ever uses OneDrive". He explained that programmers usually use Linux and not Microsoft. Working with OneDrive is very difficult if one doesn't use the Windows operating system

Some of Professor Millar's groups use tools which are more structured and powerful than a wiki but they are a little complicated for people without expertise, so he and his colleagues try to use tools that are as simple as possible.

2. Research groups need tools that meet their particular needs

In some fields, researchers need specific tools for storing, accessing and analysing data.

Dr Zielinski said that researchers usually need special tools for their data and cannot be limited to the general tools offered by the University. The "data intensive" groups either build something from scratch or customize existing systems. As an example, he mentioned BioDare, which was developed as the collaboration tool for groups working on biological clocks. It became popular outside the initial project, because rather than simply be a repository for the files, it allows data visualization and time series analysis, benefiting the users. Zielinski believes that the solution to the problem of limited funding and the need of custom solutions for data management is building on top of existing tools using their API. Existence of a write-API should be a critical criterion when selecting a product for adoption within the University.

One of the problems in developing a special tool for the project is of course lack of funding. Building them is also very time-consuming.

Dr Cahill pointed out that the bulk of research work in engineering is devoted to developing a system. He described one of his former projects where the data was shown only at the end of the process and their sole purpose was to ascertain that the tool was working. In fact, the system only served the group which was focused on testing the tool, but did not suit the other groups which were more scientific and less academic. In Cahill's opinion it is quite normal for a research group to develop its own tools, either new or based on existing ones. Scientists require tools that meet their own needs efficiently. Projects need good systems and not simply lab notebooks. He himself spent two months learning a phytoene system that one of the researchers in his study group had developed. It was a good enough product and sometimes good enough is enough. Cahill believes

that these are exactly the skills researchers need to develop so they can sell themselves later when going out into the industrial world. Skills for software developments, not just for narrow studies; software that others can evaluate and appreciate. Although publishing an academic paper is a time-consuming task, ninety percent of research time goes into developing tools for the research in his field. It's a lot of work and it also involves the protection of intellectual property. For this reason there is less open science in engineering, but he thinks this issue can be resolved.

Dr Pernet uses an international standard tool for brain scanning research where everything is done in the same place. Data files are organised inside the drive along with the metadata. It is very structured and it is easy to write programmes and pipelines for the study. Analysing and sharing the data is simplified when everyone uses the same standard. Pernet said data is simply what it is. Organising the metadata is difficult, and this is why structured tools are very important.

Professor Millar mentioned some specific structured tools for working with data that were designed specifically for biology studies. One of these is the *SEEK* software from the *FAIRDOM* collaboration, which supports the data structure standards of ISA (Investigations, Studies and Assays). This platform allows researchers to see the large picture, within a hierarchy, and how one piece of data is connected to other pieces of data. Another tool is *OpenBis* which was built by the biology community (ETH Zurich) and gives links to files and their location. One key advantage of *SEEK* is the flexibility of its permission system. The project manager decides who will have access to the metadata and to the data files (permission for individuals or groups or specific individuals inside a group). It meets the DMP requirements for international projects. Millar says there are many different systems for different purposes, each with its own functions and none that support all of a researcher's needs. Maybe in ten years there will be an all-inclusive system but right now researchers have to manoeuvre among several, different systems.

Dr Zielinski and Professor Millar said that some groups in their faculty use a tool called OMERO for dealing with microscope images. This is an open source tool specifically designed for images and has become standard. They have developed a code enabling them to transfer data easily between OMERO instances. The school agreed to pay for a pilot to host OMERO for public dissemination. According to Zielinski, OMERO usability is superior to that of DataShare for microscopy data. For example with OMERO one can browses the images collections without a need to download it all. Furthermore, there are technical metadata which are being extracted by OMERO which would either be lost in DataShare or have to be manually transferred to a file. Nevertheless, although OMERO gives added value to microscopy-based researchers, it is not suitable for other biological data.

3. Tools should offer features for analysing the data

The main part of the research process is analysing the data. In many cases researchers need special tools for data analysis.

According to Dr Zielinski, scientists use tools principally for data analysis. For example, the group working on the biological clocks created a data repository that offered them a better way to view and process their data than Excel. Most of the partners used it only for data analysis, preferring to exchange Excel files by emails for any data that did not benefited directly from the processing features. In another project, OpenNGS database helped solve specific problems with linking data to constantly changing processing scripts. While the tool is being used to help in ongoing research work, it can also transfer datasets to DataShare with one click.

4. Need to improve promotion and visibility of RDM services

Wenlong Li believes that Edinburgh University did provide information about tools and services when he joined, but he had not taken advantage of this at the time because of information overload. He thinks that the university should present the pros and cons of using its tools compared with other more popular tools. The campaign should be repetitive. He acknowledged the university's success in designing leaflets outlining their services, and the detailed guidance they offer to faculty members which helps to reduce cognitive effort.

Dr Cahill assumes that there are many accessible tools; more than he knows about. Comparing Edinburgh University's tools and services to other places where he had worked, he said they had larger budgets and more highly-skilled personnel and human resources. The RDM support team deals with data and ethics very professionally and is much more advanced than in his former places of work. The staff actively promotes its services, but these should be more connected to the academic world. He feels that researchers are too preoccupied with their own research and do not know how to cooperate. He also thinks that the RDM support group should approach PIs for the system to work well. In his experience, people fall in line with the leader's decision, whereas they wouldn't necessarily follow decisions taken by a technician or junior researcher. Leadership is needed here; it can determine how the system will work. Similarly, if the PI is satisfied with a tool, the group will adopt it even if it involves greater effort.

Conclusion

Researchers appreciate tools and services that the institution offers but are not aware of all the possibilities. It may be useful to bring the information to young researchers' attention more often and not just at the beginning of the year when they are overwhelmed with information, and to focus on encouraging the projects' PI's to use institutional tools and services. It may also be useful to emphasize the advantages of institutional tools and services compared to commercial tools (e.g. security and privacy). However, tools must be simple and user friendly for researchers to use them. Perhaps a budget for improving usability should be considered, or implementing user-friendly commercial tools in the university.

References

1. Capra, R., Vardell, E., & Brennan, K. (2014). File synchronization and sharing: User practices and challenges. *Proceedings of the American Society for Information Science and Technology*, *51*(1), 1-10.

2. Rader, E. (2009, April). Yours, mine and (not) ours: social influences on group information repositories. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2095-2098). ACM.

3. Renwick, S., Winter, M., & Gill, M. (2017). Managing research data at an academic library in a developing country. *IFLA Journal*, 43(1), 51-64.

4. Alexogiannopoulos, E., McKenney, S., & Pickton, M. (2010). *Research Data Management Project: a DAF investigation of research data management practices at The University of Northampton*. University of Northampton.

5. Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, *36*(2), 84-90.

6. Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2), 5-26.

Disclaimer: The stated opinions in this report are those of the research participants, not the Research Data Support team.

Acknowledgement: the author's month-long visit to Edinburgh was supported by The Israel Academy of Sciences and Humanities.

Appendices

Appendix 1: List of interviewees

Dr Tomasz Zielinski

Holly Tibble

Wenlong Li

Dr Brian Cahill

Dr Cyril Pernet

Professor Andrew Millar

All consented to be recorded and have quotes attributed to them in this report.

Appendix 2: Interview questions

1. What are the tools and services you use for collaboration and why did you choose them?
2. What are the differences between documents and data files?
3. What are the differences when collaborating with internal and external partners?
4. What are the considerations for collaboration when writing a Data Management Plan?
5. How do you deal with version?
6. How do you deal with security and privacy issues?